Received: 1 April 2018 | Accepted: 28 February 2019

DOI: 10.1111/2041-210X.13172

RESEARCH ARTICLE

Methods in Ecology and Evolution BRITISH ECOLOGICAL SOCIETY

A novel biomechanical approach for animal behaviour recognition using accelerometers

Pritish Chakravarty[1] | Gabriele Cozzi[2,3] | Arpat Ozgul[2,3] | Kamiar Aminian[1]

[1]Interfaculty Institute of Bioengineering (IBI-STI), Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

[2]Institute of Evolutionary Biology and Environmental Studies, University of Zurich, Zurich, Switzerland

[3]Kalahari Research Centre, Kuruman River Reserve, Van Zylsrus, South Africa

Correspondence
Pritish Chakravarty
Email: pritish.chakravarty@epfl.ch

Handling Editor: Robert B. O'Hara

Abstract

1. Data from animal-borne inertial sensors are widely used to investigate several aspects of an animal's life, such as energy expenditure, daily activity patterns and behaviour. Accelerometer data used in conjunction with machine learning algorithms have been the tool of choice for characterising animal behaviour. Although machine learning models perform reasonably well, they may not rely on meaningful features, nor lend themselves to physical interpretation of the classification rules. This lack of interpretability and control over classification outcomes is of particular concern where different behaviours have different frequency of occurrence and duration, as in most natural systems, and calls for the development of alternative methods. Biomechanical approaches to human activity classification could overcome these shortcomings, yet their full potential remains untapped for animal studies.
2. We propose a general framework for behaviour recognition using accelerometers, and develop a hybrid model where (a) biomechanical features characterise movement dynamics, and (b) a node-based hierarchical classification scheme employs simple machine learning algorithms at each node to find feature-value thresholds separating different behaviours. Using triaxial accelerometer data collected on 10 wild Kalahari meerkats, and annotated video recordings of each individual as groundtruth, this hybrid model was validated in three scenarios: (a) when each behaviour was equally represented (EQDIST), (b) when naturally imbalanced datasets were considered (STRAT) and (c) when data from new individuals were considered (LOIO).
3. A linear-kernel Support Vector Machine at each node of our classification scheme yielded an overall accuracy of >95% for each scenario. Our hybrid approach had a 2.7% better average overall accuracy than top-performing classical machine learning approaches. Further, we showed that not all models with high overall accuracy returned accurate behaviour-specific performance, and good performance during EQDIST did not always generalise to STRAT and LOIO.
4. Our hybrid model took advantage of robust machine learning algorithms for automatically estimating decision boundaries between behavioural classes. This not only achieved high classification performance but also permitted biomechanical interpretation of classification outcomes. The framework presented here provides the

flexibility to adapt models to required levels of behavioural resolution, and has the potential to facilitate meaningful model sharing between studies.

KEYWORDS
accelerometer, animal behaviour recognition, biomechanics, machine learning, meerkat, movement intensity, movement periodicity, posture

1 | INTRODUCTION

An in-depth understanding of wild animal behaviour and movement has assumed prime importance in recent years in light of an urgent need to augment our forecasting, conservation and management capacities in the face of rapid environmental change. Since gathering data on wild animals in their natural habitats is often precluded by logistical difficulty, animal-borne sensor systems that offer the possibility of continuously and remotely recording data as the animal goes about its daily life have been developed (Ropert-Coudert & Wilson, 2005). Over the two decades since their first application to remote monitoring of animal behaviour (e.g. Sellers, Varley, & Waters, 1998), animal-borne accelerometers have been employed to gain insights into the life histories of species as far apart in size, type of habitat, speed and mode of locomotion as chipmunks (*Tamias alpinus*) (Hammond, Springthorpe, Walsh, & Berg-Kirkpatrick, 2016), seals (*Leptonychotes weddellii*) (Naito, Bornemann, Takahashi, McIntyre, & Plötz, 2010), African wild dogs (*Lycaon pictus*) and cheetahs (*Acinonyx jubatus*) (Cozzi et al., 2012), vultures (*Gyps fulvus*) (Nathan et al., 2012) and blue whales (*Balaenoptera musculus*) (Goldbogen et al., 2011). To keep pace with the increasing ubiquity of accelerometry, developing methods applicable across species has become essential. For instance, general methods have been developed to infer animal energy expenditure from recorded acceleration (Wilson et al., 2006). However, despite progress in recognising animal behaviour from recorded acceleration, there is as yet no single technique that combines under one framework the virtues of easy-to-implement machine learning on the one hand, and the interpretability and robustness of biomechanically defined classification rules on the other.

Machine learning, often rather unenviably referred to as a black box (e.g. McClune et al., 2014), has been widely employed to infer animal behaviour from raw acceleration data (Nathan et al., 2012; Grünewälder et al., 2012; Gao, Campbell, Bidder, & Hunter, 2013; Bidder et al., 2014; Resheff, Rotics, Harel, Spiegel, & Nathan, 2014; Wang et al., 2015). Despite the power and ease of use of machine learning approaches, the need to develop behaviour recognition tools generalisable across species has brought to light the importance of being able to physically interpret classification rules, even at the cost of small gains in classification accuracy (Nathan et al., 2012). Model interpretability becomes particularly significant when accurately recognising relatively rarer behaviours is important. When machine learning approaches are applied to datasets that are heavily skewed in the frequency and duration of different behaviours (Watanabe, Izawa, Kato, Ropert-Coudert, & Naito, 2005; Grünewälder et al., 2012; Resheff et al., 2014; Wang et al., 2015), large overall accuracies may be obtained even when the recognition accuracy of under-represented behaviours is poor (He & Garcia, 2009); the issue of imbalanced classes has even been called a 'curse' in machine learning literature (Lemaitre, Nogueira, & Aridas, 2017). This occurs because machine learning algorithms typically seek to minimise the overall misclassification rate, and thus tend to optimise for the most frequent behaviours since they contribute most to the overall accuracy—the classification performance of rare behaviours may nevertheless remain poor. Examination of the classification rules would help understand how to improve the recognition of rarer behaviours. However, the use of many features, which is typical with machine learning approaches (>15: Nathan et al., 2012; Resheff et al., 2014; Wang et al., 2015), implies that the resulting classification rules are a function of that many variables, and deciphering these high-dimensional complex rules and pinpointing the reasons behind misclassification of under-represented (rarer) but important behaviours quickly becomes intractable.

Confining the role of machine learning algorithms to threshold-finding within a classification scheme predefined on the basis of biomechanical knowledge of animal movement dynamics, and using biomechanically significant movement descriptors (features) within the scheme, may help solve these difficulties. Such an approach is made possible by the fact that there are certain natural commonalities in the movement biomechanics of any animal: different behaviours may be carried out more or less intensely, in characteristic postures, erratically or in a periodic manner. Even though the actual names for different animal behaviours will depend on the environment and animal's context, broad behavioural categories will still lend themselves to description through the biomechanical principles of intensity, posture and periodicity.

Several studies have employed biomechanical considerations to characterise behaviours central to an animal's life, such as foraging, locomotion and resting (e.g. Shepard et al. 2008); in these studies, a majority of the acceleration-derived metrics used for behaviour recognition were aimed at quantifying the three biomechanical principles of intensity, posture and periodicity. For instance, one study that categorised acceleration data in Adélie penguins (*Pygoscelis adeliae*) (Yoda et al., 2001) considered the division of behavioural classes into static and dynamic categories, discriminated between static behaviours based on posture and identified locomotion as being periodic. The use of periodicity metrics may help tease apart even highly context-, environment- and mode-of-locomotion-dependent behaviours such as locomotion and foraging. For instance, periodicity metrics have since been used to characterise locomotor movement across avian (Laich, Wilson, Quintana, & Shepard, 2008), marine (Kawabe et al. 2003) and terrestrial (Soltis et al., 2012)

species; the aperiodicity of foraging has been alluded to for Little Penguins (*Eudyptula minor*) diving for unpredictably located and mobile prey (Ropert-Coudert, Kato, Wilson, & Cannell, 2006; Watanabe & Takahashi, 2013). Despite extensive use of such biomechanically significant parameters in their behaviour recognition schemes, previous methods leave room for improvement—parameter thresholds for building classification rules were manually chosen (Laich et al. 2008), and training data used for creating the model came only from a few captive animals (McClune et al., 2014; Soltis et al., 2012).

In this study, we present a general behaviour recognition framework in the form of a hybrid model that combines general biomechanical principles on the one hand, and machine learning tools on the other. The proposed hybrid model consists of a tree-like classification framework predefined on the basis of biomechanical considerations, where specific combinations of acceleration-derived biomechanical descriptors capturing movement biomechanics of behavioural categories studied across a range of species are used in conjunction with robust machine learning algorithms at each node of the tree. The use of biomechanics to instruct the classification makes the model interpretable, and the use of machine learning at each node of the tree completely automates the search for optimal metric thresholds separating different behaviours. We showcase the application and benefits of this hybrid behaviour recognition framework on data collected on wild meerkats (*Suricata suricatta*), where the classification of their main activities such as vigilance, foraging, resting and running is essential for characterising their social and individual behaviour. We validated this hybrid model against synchronised, annotated video camera footage considered as the groundtruth, and compared its performance with benchmark measures obtained with alternative classification methods based entirely on classical machine learning.

2 | MATERIAL AND METHODS

2.1 | Biomechanically driven behaviour recognition and validation

2.1.1 | Quantifying movement biomechanics through acceleration-derived features

Raw triaxial acceleration data may be summarised in the form of quantifiable biomechanical descriptors of movement, which can then be used as features capable of discriminating between different behavioural categories. The features to be developed will be strongly dependent on the desired ethogram; here we shall consider an archetypal, general ethogram consisting of three common behaviours—locomotion, resting and foraging. We focused on three biomechanical principles to characterise these behaviours: posture, movement intensity and periodicity. These principles have been previously used to discriminate between behaviours for a range of species; we aim to synthesise and combine these existing but scattered principles within one biomechanically defined classification framework. We quantify these three descriptors for each fixed-duration sliding window *w* containing *N* triaxial acceleration values recorded along the surge (a_{surge}), sway (a_{sway}) and heave (a_{heave}) axes as follows.

Posture

Common measures of posture involve estimating how the sensor is oriented with respect to Earth's gravity. Computation is done by averaging acceleration recorded along each axis of the accelerometer, assuming that acceleration due to dynamic bodily movement shows up as oscillations around a constant, static value corresponding to Earth's gravity (e.g. Yoda et al., 2001). For instance, the mean of surge acceleration in window *w*, $a_{surge,w}$, may be computed as:

$$meanSurge_w = \frac{\sum_N \alpha_{surge,w}}{N} \quad (1)$$

Intensity

Contrary to posture estimation, movement intensity is commonly characterised by quantifying dynamic acceleration, which is obtained by removing static acceleration corresponding to Earth's gravity from total recorded acceleration. For this, we considered the use of $stdNorm_w$, the standard deviation of the Euclidean norm of the triaxial acceleration vector ($|\vec{a}_w|$), which is equivalent to computing Vectorial Dynamic Body Acceleration (VeDBA, McGregor, Busa, Yaggie, & Bollt, 2009):

$$stdNorm_w = std\left(|\vec{a}_w|\right) \quad (2)$$

Periodicity

To quantify movement periodicity, we analysed the frequency content of the signal through the use of the Fourier transform (FT). The FT of aperiodic signals such as acceleration recorded during foraging will be relatively flat, whereas that computed for periodic signals such as those recorded during locomotion will be marked by the presence of a clear, tall peak, usually at a characteristic frequency. This peak height was considered as a measure of signal periodicity, and may be computed as follows. For each window *w*, the acceleration signal from each of the three axes recorded at a sampling frequency of F_s may be: (a) normalised (zero mean and unit energy); (b) low-pass filtered; (c) zero-padded and windowed; and (d) FT-ed with a resolution of *U* Hz (corresponding to FT computation at $L = F_s/U$ equally spaced frequencies between 0 and F_s). From the FT of acceleration along each axis, the square of the magnitude of each Fourier coefficient ($c^2_{f_i,w,sway}, c^2_{f_i,w,surge}, c^2_{f_i,w,heave}, i \in 1 \ldots L$), corresponding to the power of the signal at frequency f_i, may be computed and averaged over the three axes. Finally, from this axis-averaged FT, the maximum power obtained across all frequencies $f_i\ (i \in 1 \ldots L)$ may be chosen as a measure of the periodicity of the signal. This feature will henceforth be referred to as *fftPeakPowerAvg*.

$$fftPeakPowerAvg_w = \max_L \frac{c^2_{f_i,w,sway} + c^2_{f_i,w,surge} + c^2_{f_i,w,heave}}{3} \quad (3)$$

Note that actual values of the signal processing parameters, such as type, order and cut-off frequency of the low-pass filter, length of zero-padding, type of window and *U*, will depend on the characteristics of the recorded acceleration signal, such as signal bandwidth and sampling frequency.

2.1.2 | A biomechanically defined hierarchical classification scheme with automated feature-threshold computation

The proposed approach involves predefining a hierarchical tree-like scheme that classifies broader behavioural categories into increasingly specific ones up to the desired level of behavioural resolution. Each node of this tree uses one or more features tailored to the classification at that node. A predefined hierarchical scheme has two interesting characteristics that make it more advantageous over the common classical machine learning approach of directly classifying behaviours at their highest resolution. The first advantage concerns the ease of dealing with imbalanced classes. It is quite probable that when specific behaviours are grouped into compound categories at the higher placed nodes of the tree, a relatively rarer behaviour gets clubbed into the same compound category with a more frequent behaviour. As a result, the problem of imbalanced classes is not encountered at the higher placed node, and is deferred to a lower node where the rare behaviour can no longer be grouped into a compound category with the more frequent behaviour. Thus, if poor recognition accuracy of a rare behaviour is encountered, one can precisely identify the node at which the misclassification occurs. Moreover, since the input features are tailored to each node, one can understand which features to add or modify at the culprit node to improve classification. This process of optimising the model for each individual behaviour may be very difficult to do with machine learning approaches that classify directly up to the finest behaviour resolution level using classification rules which are a function of many (>15) features. Secondly, when higher behavioural resolution is desirable, specific behaviours themselves can be considered as compound categories and separated into finer behaviours by adding nodes below them. For instance, adding a node below 'locomotion' could distinguish between slow and fast locomotion. In classical machine learning, if behavioural classes are added retrospectively, the entire model would have to be built anew.

The precise form of the classification tree for a specific application will depend on the ethogram of interest, so will the features to be given as inputs to each node. Here we demonstrate the construction of a classification tree for the archetypal ethogram considered in Section 2.1.1 consisting of locomotion, resting and foraging. For the classification of these behaviours, the first node would separate the static (resting) and dynamic (locomotion and foraging) behavioural categories. At the second stage of this tree, one node below the dynamic category would separate the two dynamic behaviours: locomotion and foraging. In this scheme, each node divides a parent behavioural category into exactly two daughter behavioural types. To accomplish this binary classification at each node, appropriate user-chosen biomechanically significant features (cf. Section 2.1.1) may be given as inputs to commonly employed binary machine learning algorithms (such as Support Vector Machines (SVM)) to obtain optimal feature-value thresholds in a completely automated fashion. For instance, at the first node of the classification tree described here, *stdNorm*, a measure of movement intensity, may be used as a single-feature input to an SVM that will separate behaviours into the static and dynamic behavioural categories.

2.1.3 | Model validation

To test the utility of a behaviour recognition method, one needs to validate the predictions made by it against groundtruthed data—the latter typically coming from direct observation or video annotation of the behaviours of interest. Typically, a video camera synchronised with the animal-borne accelerometer is used to film the animal while the accelerometer records data; the groundtruth is then obtained by having an expert assign behaviour labels to sections of the video based on a suitably defined ethogram. This process is then repeated for multiple individuals to capture inter-individual variation in behaviours. Finally, windows of acceleration corresponding to observed behaviours of interest are extracted from the continuous acceleration data stream to obtain bouts of labelled behaviour of fixed duration. To evaluate the effect of imbalanced datasets and inter-individual variability on model classification, we discuss three different forms of cross-validation.

Equally distributed behaviour 10-fold cross-validation (EQDIST)

EQDIST evaluates model performance when the dataset has an equal number of bouts of each behaviour. It involves conducting standard 10-fold cross-validation on subsampled datasets where the sample size for each behaviour is made equal. This is done by first pooling data from all individuals together, and then randomly selecting as many bouts from each behaviour as the one with the least number of bouts.

Stratified 10-fold cross-validation (STRAT)

STRAT evaluates model performance when some behaviours may be under-represented or rarer in the dataset than others. It involves pooling data from all individuals together and then dividing the pooled dataset into 10 equal parts in such a way that the proportion of bouts from each type of behaviour in each fold is equal to that in the entire, unmodified dataset.

Leave-one-individual-out cross-validation (LOIO)

LOIO evaluates model performance when inter-individual variation is taken into account. It involves training a model using data pooled over all individuals except one, and then testing this model on data from the individual left out. This process is repeated until each individual has been the 'test' individual once.

We used confusion matrix-based metrics to evaluate and compare model performance. These performance statistics included three behaviour-specific metrics (sensitivity, precision and specificity), and overall model accuracy (see Appendix S1 for definitions).

2.2 | Case Study: Kalahari Meerkats

2.2.1 | Data collection and groundtruthing

Fieldwork was conducted at the Kalahari Meerkat Project, a long-term study of wild meerkats in the South African Kalahari Desert (Clutton-Brock, Gaynor, & McIlrath, 1999). Eleven data-recording sessions were done on 10 adult meerkats (seven males, three females, body mass

667 ± 98 grams, age 24 ± 15 months); one of the individuals was recorded twice. For each individual, data were collected over 3 hr in the morning, which corresponds to the duration of normal morning meerkat activity (meerkats become inactive as temperatures soar around the midday hours). Individuals were captured using methodology described in Jordan, Cherry & Manser, 2007 and fitted with a collar equipped with an inertial measurement unit (IMU) (adapted version of Physilog IV, GaitUp SA, Switzerland) of size 35 mm × 29 mm × 19 mm and measuring triaxial acceleration at 100 Hz/axis with a range of ± 156.96 m/s^2 (corresponding to ± 16 times the acceleration due to Earth's gravity) with 16-bit resolution. The overall collar weight was <25 g; collars of this size and weight have been shown not to affect meerkat behaviour (Golabek, Jordan, & Clutton-Brock, 2008). The collar was positioned so that the axes of the accelerometer were oriented as shown in Figure 1. The accelerometer was calibrated prior to each capture according to a standard procedure (Ferraris, Grimaldi, & Parvis, 1995). After release, the collared animal was filmed using a handheld video camera recording at 25 frames/second and synchronised electronically with the IMU system. All videos were annotated using Solomon Coder (Version: beta 17.03.22). This video annotation served as the groundtruthing for our behaviour recognition scheme.

2.2.2 | Meerkat behaviours and hierarchical classification scheme

Four different behaviours (Figure 2) were considered for the ethogram based on their biological significance:

- *Vigilance*: representative of the animal's general stress or alertness level. The meerkat is stationary and lifts its head and torso up to survey its surroundings.
- *Foraging*: can help derive proxies for body condition, which could, in turn, play a role in determining the animals' survival and social status (Hodge, Manica, Flower, & Clutton-Brock, 2008). Most meerkat prey live underground; foraging involves scanning the ground, digging, and handling and ingesting prey.
- *Running*: high-speed locomotion from one point to another with presumably high energy expenditure. Prolonged running events are rare and typically mark important events such as aggressive interactions with rival meerkat groups.
- *Resting*: periods of inactivity, mainly due to fatigue or excessive heat. Typically, the meerkat lies down with its body flat on the ground.

Video clips of each behaviour are provided as Supplementary Information. Social context-dependent meerkat behaviours such as grooming and territory marking were excluded from the ethogram.

Based on the description of the behaviours of interest here and the framework presented in Section 2.1, a three-node hierarchical scheme was devised to classify meerkat behaviour, as shown in Figure 2. Movement intensity (*stdNorm*) and posture (*meanSurge*) were used to separate static (vigilance and resting) and dynamic (foraging and running) behaviours in the first node. At the second node, posture (*meanSurge*) was used to distinguish vigilance from resting—while

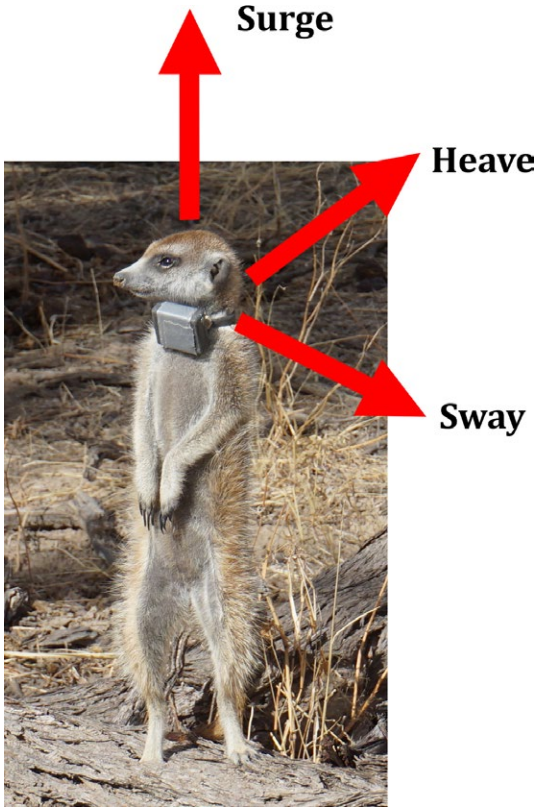

FIGURE 1 Sensor axes orientation. The image shows the animal displaying typical bipedal vigilance behaviour

the animal's torso is typically upright during vigilance, it is flat during resting. At the third node, movement intensity (*stdNorm*) and periodicity (*fftPeakPowerAvg*) were used to separate foraging from running—running was observed to involve faster and more periodic limb movements compared to foraging. At each node, a separate machine learning algorithm (M1, M2 and M3 in Figure 2) was chosen from four candidates: Naïve-Bayes (NB), Linear Discriminant Analysis (LDA), Logistic Regression (LR) and linear-kernel Support Vector Machine (SVM). We shall henceforth refer to each such M1-M2-M3 combination as one 'hybrid model'. Thus, since four candidates were possible at each of the three nodes, a total of 4 × 4 × 4 = 64 hybrid models were tested to find the best one. Scikit-learn (Pedregosa, 2011, version 0.19.0) was used to implement all machine learning models (using their default configurations) in `Python`.

2.2.3 | Feature computation

A sliding window *w* of size two seconds was considered for feature computation; this provided sufficient temporal resolution of behaviour and was short enough to capture bouts of running, the behaviour with the shortest duration. Successive windows had an overlap of 50%. Windows containing transitions between different behaviours were excluded; each window thus contained acceleration data corresponding to exactly one video-labelled behaviour.

For each window *w*, acceleration along the surge axis ($a_{surge,w}$) only was averaged ($meanSurge_w$) and used to estimate neck inclination

(Equation 1), as values along this axis are least susceptible to changes due to possible rotations of the collar around the axis of the meerkat's cylindrical neck. For *fftPeakPowerAvg* computation (cf. Section 2.1.1 under 'Periodicity'), the parameters used were: F_s = 100 Hz, U = 0.01 Hz, L = 10000. The low-pass filter was a fourth-order Butterworth filter with a cut-off frequency of 10 Hz. The length of zero-padding was 1 s, and the Blackman–Harris windowing function was used. All feature computations were done using MATLAB R2016b. MATLAB code to compute *fftPeakPowerAvg* is supplied in Supplementary Information.

2.2.4 | Model validation

For EQDIST, 10 synthetically equalised datasets were generated using the Imbalanced-learn (Lemaitre et al., 2017, version 0.3.0) package in Python. STRAT was implemented using Scikit-learn (Pedregosa, 2011). For both EQDIST and STRAT, confusion matrices obtained from each test fold were added together to produce an aggregated confusion matrix from which performance statistics were calculated. LOIO was done for individuals for which all four behaviours were recorded. Performance statistics for each individual were computed separately, and their mean and standard deviation across individuals were reported.

2.2.5 | Alternative classification methods: benchmarking against classical machine learning approaches

To benchmark our hybrid model's results against those obtained with alternative classification methods based entirely on classical machine learning, four commonly employed algorithms were considered: Naïve-Bayes (NB), K-Nearest Neighbours (KNN, with K = 5), Random Forest (RF, with 10 trees) and Support Vector Machine (SVM) with a linear kernel. Features presented in a recent review of animal behaviour recognition

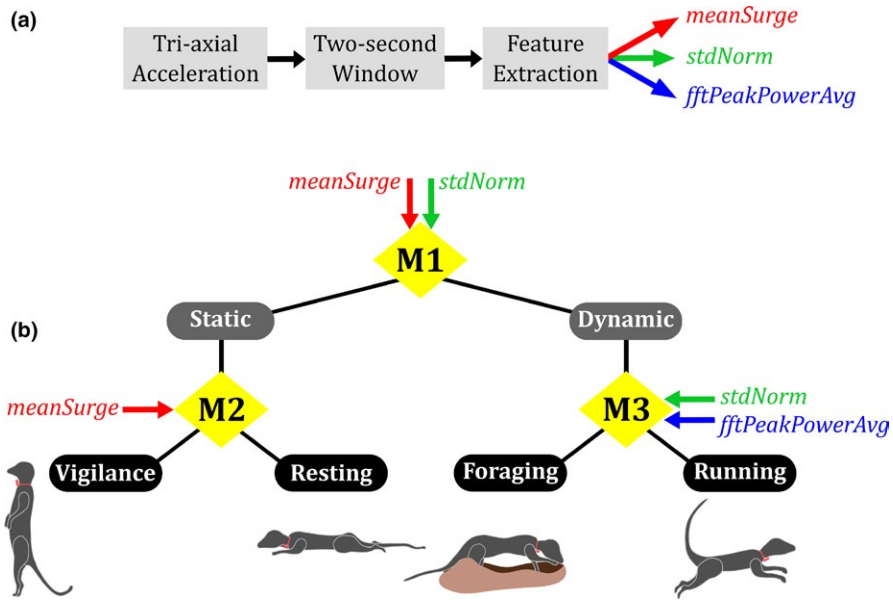

FIGURE 2 Biomechanically informed behaviour recognition scheme. (a) Workflow showing feature extraction from triaxial acceleration; (b) At each of the three nodes of the classification scheme, a separate machine learning algorithm (M1, M2, M3) is trained with specific biomechanical features encapsulating information on posture (*meanSurge*), movement intensity (*stdNorm*) and periodicity (*fftPeakPowerAvg*)

TABLE 1 Summary of data collected. Data were collected on 10 unique individuals; data from recording session numbers 4 and 7 were collected on the same individual. A bout refers to a window *w* of two seconds containing one video-labelled behaviour

Recording session number	Vigilance	Resting	Foraging	Running	Bouts per recording session
1	4,594	2,114	1,562	69	8,339
2	3,896	120	5,315	29	9,360
3	1,453	0	6,278	38	7,769
4	5,221	0	2,823	161	8,205
5	1,890	0	6,134	169	8,193
6	1,639	744	4,438	98	6,919
7	4,785	156	3,498	40	8,479
8	71	0	4,841	20	4,932
9	4,283	0	1,713	43	6,039
10	1,906	0	4,407	84	6,397
11	1,782	661	5,398	77	7,918
Bouts per activity	31,520	3,795	4,6407	828	82,550 (total bouts)

(Nathan et al., 2012) were considered as input to these machine learning algorithms. They considered statistical features—mean, standard deviation, skewness, kurtosis, maximum and minimum—computed from data from each accelerometer axis as well as the acceleration norm. For consistency, we retained only those features that either used only the surge axis or all three axes together—16 such features were identified. Further, to meaningfully compare results with our three-feature hybrid model, we selected the top three features from this set of 16, thereby ensuring that both approaches would have the same complexity in terms of feature-space dimensionality. Further details on feature selection can be found in Appendix S2.

3 | RESULTS

3.1 | Observed behaviour

A total of 105,604 2-s bouts of video-labelled behaviour were collected. About 12.3% of these bouts contained transitions from one observed behaviour to another, in 6.2% of them the animal was not caught on camera and 3.3% contained social behaviour such as grooming (cf. Section 2.2.2)—these bouts were excluded from the dataset. The remaining 82550 bouts corresponded to the four behaviours of interest, as presented in Table 1. Foraging (56.2% of retained bouts) and vigilance (38.2%) were the most common behaviours, whereas running (1%) was the rarest. Typical signals recorded for each behaviour are shown in Figure 3.

3.2 | Performance evaluation

Out of the 64 possible combinations for the M1-M2-M3 hybrid model, we found that SVM-SVM-SVM performed the best across all three cross-validation methods. The linear-kernel SVM not only fully automated the search for robust feature-value thresholds but also yielded simple linear decision boundaries, thereby rendering classification rules transparent and intuitive, as shown in Figure 4. The next three subsections give details on the performance of the SVM-SVM-SVM hybrid model for each cross-validation method tested, and benchmark them against results obtained with classical machine learning using the same number of features. Results obtained with classical machine learning using all 16 features are provided in Appendix S3.

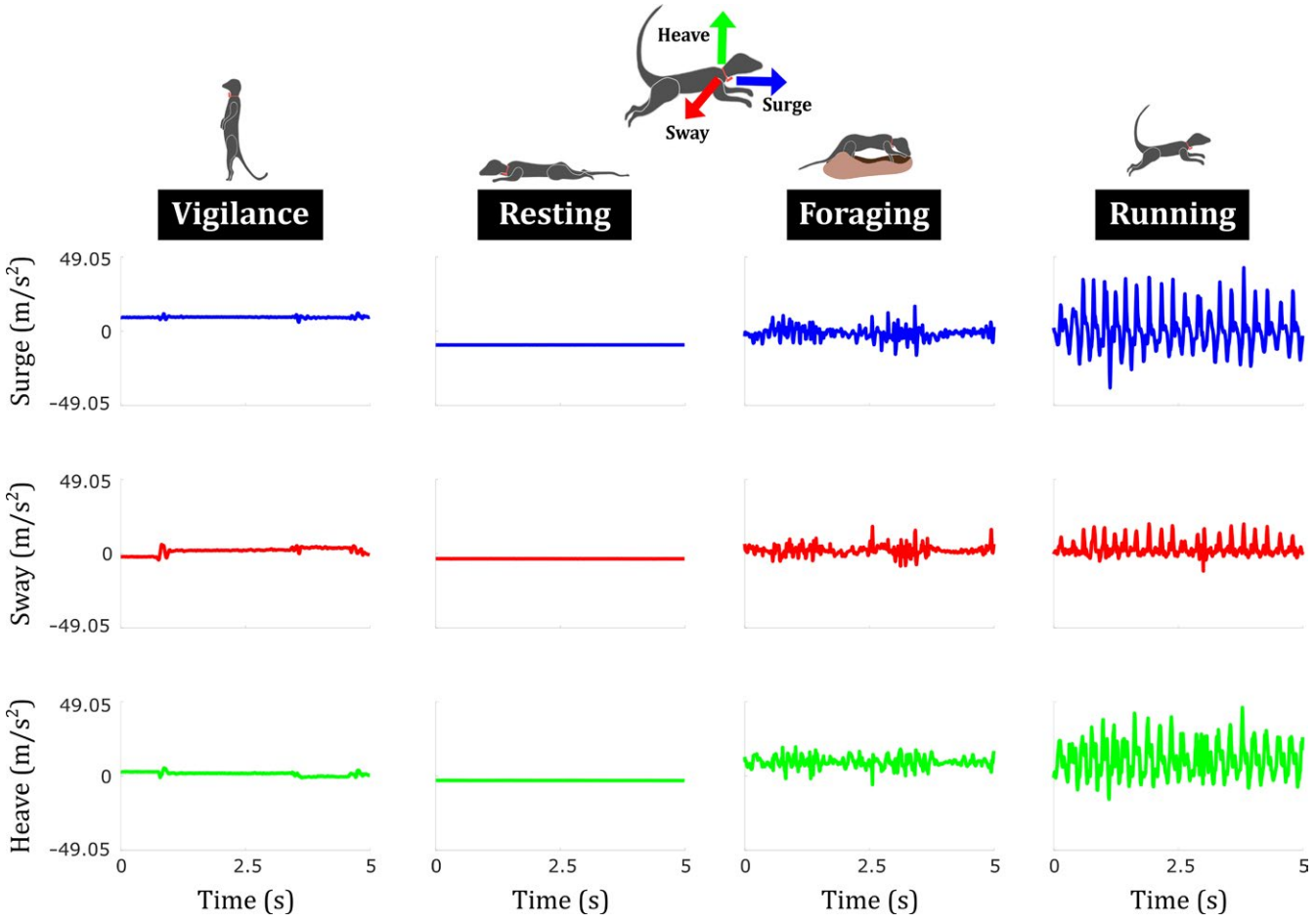

FIGURE 3 Five-second snapshots of the four behaviours of interest for a typical individual (dataset #1). During vigilance, the meerkat is still; short perturbations in the signal occur when the head turns to scan the surroundings. During resting, the meerkat remains still and the surge signal has a different intercept compared to that during vigilance. Foraging produces erratically varying signals due to site-dependent digging and manoeuvring. Running is a fast, rhythmic activity marked by a highly periodic signal

3.2.1 | EQDIST

Results for EQDIST in Table 2 show that for behaviour-specific metrics, the best hybrid model (SVM-SVM-SVM) performed better in terms of sensitivity (average 3.5% across the four behaviours), precision (average 3.3%) and specificity (average 1.1%) compared to the best machine learning model (K-Nearest Neighbours).

3.2.2 | STRAT

Results for STRAT in Table 3 show that for behaviour-specific metrics, the best hybrid model (SVM-SVM-SVM) performed better in terms of sensitivity (average 5.1% across the four behaviours), precision (average 3.1%) and specificity (average 0.7%) compared to the best machine learning model (K-Nearest Neighbours).

3.2.3 | LOIO

For LOIO, data from recording sessions numbered 3, 4, 5, 8, 9 and 10 (see Table 1) were discarded since they did not contain any resting behaviour. The mean and standard deviation of all performance metrics for LOIO with data from the five retained recording sessions (corresponding to five unique individuals) are shown in Table 4. For behaviour-specific metrics, the best hybrid model (SVM-SVM-SVM) performed better in terms of mean sensitivity (average 1.1% across the four behaviours), mean precision (average 4.3%) and mean specificity (average 1.3%) compared to the best machine learning model (Random Forest).

4 | DISCUSSION

We presented a physically intuitive behaviour recognition framework based on a hybrid model that combines movement biomechanics and robust machine learning. We showed that with our biomechanically defined node-based hierarchical classification framework, acceleration-derived features summarising movement biomechanics could be used in conjunction with existing machine learning algorithms to recognise behaviour from triaxial acceleration data. Using data collected on 10 wild meerkats, we demonstrated the efficacy of our hybrid model in scenarios where one or more behaviours are rarer or under-represented in the dataset compared to others, and when there is

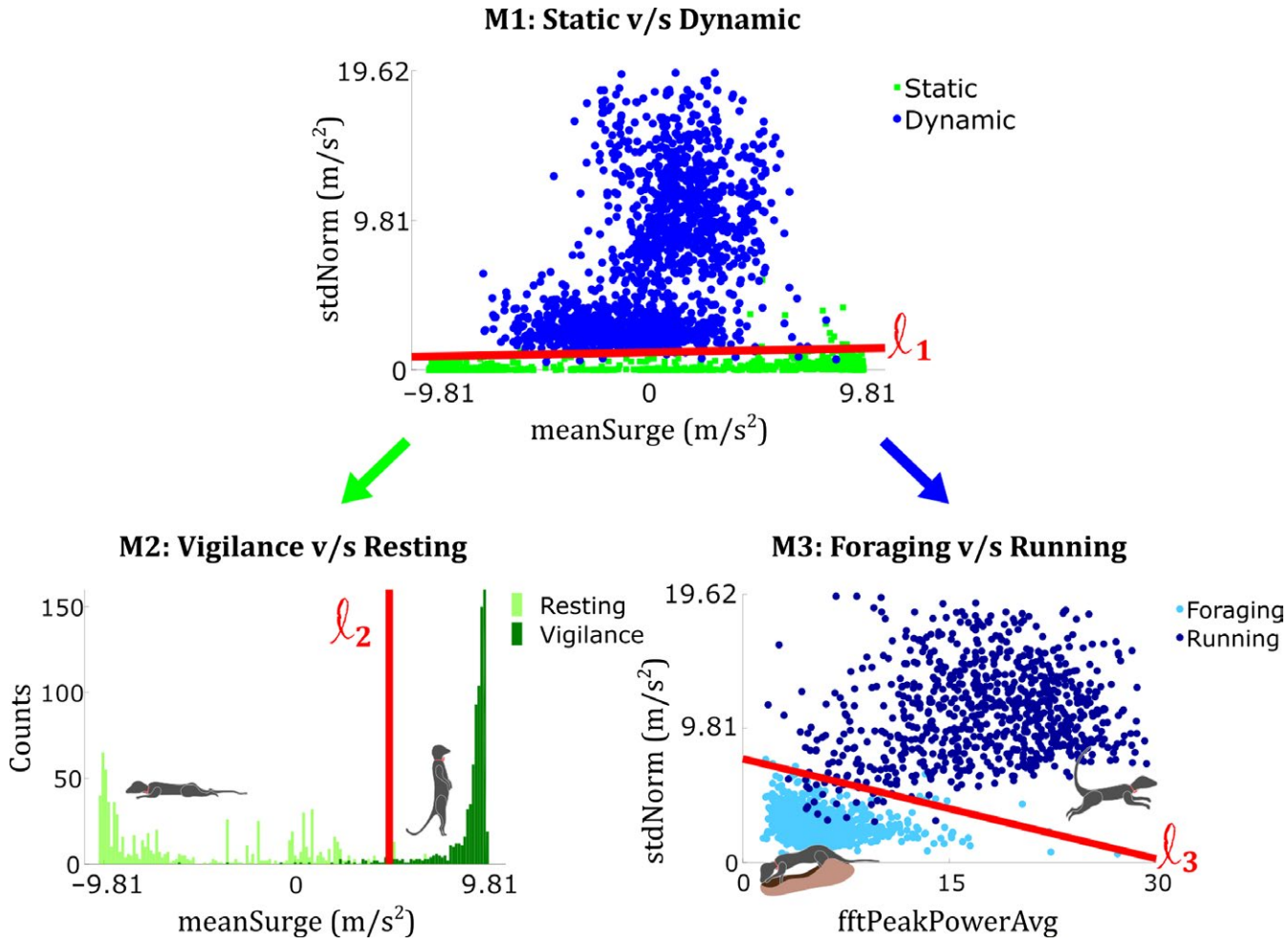

FIGURE 4 Transparent classification. Linear decision boundaries (ℓ_1, ℓ_2 and ℓ_3) found by the SVM-SVM-SVM hybrid model (for data used during EQDIST) render the classification scheme transparent and physically intuitive

TABLE 2 EQDIST cross-validation results. The performance of the SVM-SVM-SVM hybrid model is benchmarked against that of classical machine learning

Approach	Model	Vigilance			Resting			Foraging			Running			Overall accuracy (%)
		Sen. (%)	Spec. (%)	Prec. (%)	Sen. (%)	Spec. (%)	Prec. (%)	Sen. (%)	Spec. (%)	Prec. (%)	Sen. (%)	Spec. (%)	Prec. (%)	
Hybrid	SVM-SVM-SVM	95.2	98.2	94.7	93.7	98.6	95.8	97	98	94.3	96.7	99.3	97.9	95.7
Classical Machine Learning	NB	43.5	96.7	81.6	88.9	82.3	62.6	95.7	97.5	92.6	97.8	98.8	96.5	81.5
	KNN	91.3	94.5	84.6	82.7	97.9	93	96.9	97.9	93.9	97.9	99.3	97.9	92.2
	RF	90.5	94.4	84.3	83.3	97.5	91.8	96.7	98	94.2	97.5	99.4	98.2	92
	SVM	53.6	96.2	82.3	87.6	85.7	67.1	97.3	97.5	92.8	97.7	99.4	98.2	84

NB, Naïve-Bayes; KNN, K-Nearest Neighbours; RF, Random Forest; SVM, Support Vector Machine.

TABLE 3 STRAT cross-validation results. The performance of the SVM-SVM-SVM hybrid model is benchmarked against that of classical machine learning

Approach	Model	Vigilance			Resting			Foraging			Running			Overall accuracy (%)
		Sen. (%)	Spec. (%)	Prec. (%)	Sen. (%)	Spec. (%)	Prec. (%)	Sen. (%)	Spec. (%)	Prec. (%)	Sen. (%)	Spec. (%)	Prec. (%)	
Hybrid	SVM-SVM-SVM	97.1	98.8	98.1	85	99.4	87.1	99.3	97.8	98.3	85.9	99.9	92.1	97.7
Classical Machine Learning	NB	96.4	92	88.2	0	100	0	98	96.6	97.3	94.7	99.3	58.1	92.9
	KNN	95.8	96.6	94.5	64.1	99.3	81.4	98.8	97.4	98	88.4	99.9	89.4	96
	RF	95.7	96.5	94.4	64.7	99.2	80.2	98.8	97.5	98.1	86.7	99.9	90.7	95.9
	SVM	96.8	92.1	88.4	0	100	0	99.2	96.6	97.4	82.6	99.9	91.3	93.6

NB, Naïve-Bayes; KNN, K-Nearest Neighbours; RF, Random Forest; SVM, Support Vector Machine.

TABLE 4 LOIO cross-validation results. The performance of the SVM-SVM-SVM hybrid model is benchmarked against that of classical machine learning. Performance metrics were calculated separately for each test individual and their mean and standard deviation across test individuals are shown here

Approach	Model	Vigilance			Resting			Foraging			Running			Overall accuracy (%)
		Sen. (%)	Spec. (%)	Prec. (%)	Sen. (%)	Spec. (%)	Prec. (%)	Sen. (%)	Spec. (%)	Prec. (%)	Sen. (%)	Spec. (%)	Prec. (%)	
Hybrid	SVM-SVM-SVM	95.8 ± 2.8	98.4 ± 1.2	96.4 ± 4.5	71.4 ± 23.6	98.9 ± 1.2	81.1 ± 28.0	98.8 ± 1.0	97.4 ± 1.5	95.3 ± 7.0	86.3 ± 13.2	99.9 ± 0.1	89.1 ± 11.1	96.5 ± 1.8
Classical Machine Learning	NB	76.8 ± 22.4	82.6 ± 22.3	78.7 ± 14.4	31.7 ± 45.5	90.9 ± 12.7	2.1 ± 2.9	97.3 ± 1.0	96.1 ± 1.8	94.4 ± 6.6	98.6 ± 2.0	99.3 ± 0.8	61.9 ± 20.6	78.9 ± 8.2
	KNN	92.5 ± 2.6	93.1 ± 6.2	88.8 ± 8.6	62.6 ± 18.4	98.3 ± 0.8	63.4 ± 29.8	98.2 ± 0.9	97.4 ± 1.1	96.2 ± 4.5	90.2 ± 12.5	99.9 ± 0.2	90.7 ± 10.1	93.1 ± 2.8
	RF	92.7 ± 3.2	93.7 ± 5.5	89.7 ± 7.3	68.5 ± 17.8	98.4 ± 1.0	66.9 ± 28.4	98.0 ± 1.0	97.6 ± 1.1	96.6 ± 3.8	88.8 ± 12.0	99.9 ± 0.2	91.4 ± 11.7	93.6 ± 2.4
	SVM	95.5 ± 2.6	82.0 ± 21.8	79.3 ± 14.8	0	100	0	98.7 ± 1.0	96.1 ± 1.8	94.5 ± 6.5	84.8 ± 12.0	99.9 ± 0.1	90.8 ± 9.4	88.2 ± 10.5

NB, Naïve-Bayes; KNN, K-Nearest Neighbours; RF, Random Forest; SVM, Support Vector Machine.

inter-individual variability. We showed that our hybrid model's predictions were more accurate than those obtained with alternative classification methods based on classical machine learning using the same number of features.

Our results shed light on the complex interplay of several factors involved in behaviour recognition: feature selection, class imbalance, class separability, model selection and persistent misclassification. While the feature selection methods used in classical machine learning found measures of movement intensity (*stdNorm*, *stdSurge*; Appendix S2), they failed to select a measure of posture. This proved to be especially problematic for distinguishing the two static behaviours, vigilance and resting, since they primarily differ in posture. Consistently poorer resting and vigilance detection performance resulted when a measure of posture was excluded (classical machine learning) compared to when it was included (*meanSurge* in the hybrid model) even in EQDIST, where matters are not yet complicated by class imbalance. The effect of not including a measure of posture gets amplified when naturally occurring class imbalance is introduced in STRAT: the classical machine learning models NB and SVM completely miss the much rarer resting behaviour (vigilance bouts outnumber resting bouts 8.3:1), whereas for KNN and RF, resting detection sensitivity plummets to below 65% and precision to below 82%. Model response to class imbalance may vary unpredictably: classical machine learning with KNN and RF yields fair performance across EQDIST, STRAT and LOIO; SVM, on the other hand, first recognises resting during EQDIST (albeit poorly) despite the absence of a feature describing posture, then completely misses resting behaviour during STRAT and LOIO (even though overall model accuracy still remains high: 93.6% and 88.2% respectively), and then performs well when all 16 features are included (Appendix S3). This indicates the difficulty of generalising model behaviour across datasets when feature choice is left to an automatic algorithm. However, the hybrid model, where the SVM algorithm was only used to find feature-value thresholds within a biomechanically predefined tree-like classification structure with task-specific features (i.e. biomechanically appropriate features for each node in the classification tree), consistently performs well across EQDIST, STRAT and LOIO. Class imbalance may be less problematic when class separability is high. Running, despite being the rarest class in our dataset (foraging instances outnumber running instances 56:1) is recognised fairly accurately across EQDIST, STRAT and LOIO for the KNN, RF and SVM classical machine learning models, and the hybrid model. This might be because measures of intensity (*stdNorm*, *stdSurge*) were particularly effective at separating foraging and running: this can be seen from Figure 4, where the magnitude of <1 of the slope of the decision boundary ℓ_3 implies that the feature on the vertical axis (*stdNorm*) is more discriminating than the one on the horizontal axis (*fftPeakPowerAvg*).

In addition to exponentially greater model complexity, perhaps the greatest disadvantage of increasing the number of features in classical machine learning is the loss of ability to understand why misclassifications persist. The performance of classical machine learning becomes comparable to that of the hybrid model when all 16 features are used (Appendix S3), yet even with a fivefold increase in the number

of features (from 3 to 16), resting detection performance still remains poor. Examination of the classification rules to understand the source of misclassification is rendered unmanageable by the fact that, for 16 features, 16-dimensional space will need to be analysed. With the hybrid model, however, due to the hierarchical nature of the classification scheme, it is easy to pinpoint where and why resting misclassification occurs. The rarer resting behaviour is clubbed with a frequent behaviour, vigilance, into the static behavioural category, and the rarer running behaviour is clubbed with a frequent behaviour, foraging, (foraging bouts outnumber running bouts 56:1) into the dynamic category. Thereafter, since the static and dynamic behavioural categories are separated accurately (Table S7), it is easy to see that the primary deterioration of resting recognition accuracy must occur at the M2 node. This node employs an SVM which uses only one feature as input, *meanSurge* (measure of posture)—this thus indicated that there may have been limitations to our hypothesis about posture during resting and/or vigilance. Indeed, re-consulting the groundtruthing video data, we discovered that the major source of erroneous resting detection was the disproportionately large contribution of an additional unexpected curled-up resting posture (different from the belly-flat position typically observed) of a single individual (see Appendix S4). Thus, new insight into the groundtruthing data itself was obtained because of the interpretability offered by our hybrid model; this is in contrast to machine learning approaches, which have to rely completely on groundtruthing data.

Our behaviour recognition framework offers two other key advantages over existing methods. Firstly, in our hybrid model, robust machine learning algorithms were tasked with searching for feature-value thresholds, thus making the mechanism of decision boundary selection automatic, general, clear and easy to implement. In the tree-based classification model presented by McClune et al., 2014, analyses were based on data from a single, tame individual, and it was not clear how their iterative feature-threshold selection method could be extended to data from more individuals. Secondly, our classification scheme has the potential to enable meaningful model sharing across studies by offering the added advantage of being adaptable to desired levels of behavioural resolution. For studies where coarse behavioural resolution is sufficient, our hybrid model can be used to simply determine when the animal was static or dynamic. For studies requiring higher behavioural resolution, our scheme may be used as a template upon which new nodes, possibly using additional biomechanically derived features, may be added further down the tree to accommodate new behaviours when needed. For the meerkat ethogram chosen for this study, three biomechanically interpretable features proved to be enough but, for instance, if one were interested in characterising meerkat foraging effort, one could add an additional node below 'foraging' and split it into 'digging' and 'non-digging' through a peak-detection-based metric indicating front-paw swipes made during digging. In a classical machine learning approach, the whole model would need to be rebuilt from scratch if new behaviours were to be added; in our approach, adding higher behavioural resolution to a given 'coarser' model will not impact the existing model parameters. This could enable separate studies on the same species to add upon a single model until the required behavioural resolution has been achieved. This high-resolution model may then be made available for future studies on that species.

Choosing appropriate sensor parameters may be a crucial component of achieving accurate behaviour classification. Even though Gao et al., 2013 followed a hierarchical classification scheme employing SVMs, the web-based system they developed limited the input sampling frequency to only 1 Hz which, they found for some species, was insufficient to extract meaningful information from the frequency-domain features they used. In our meerkat study, requisite signal processing techniques employed to compute movement periodicity through *fftPeakPowerAvg* might have played an important role in successfully distinguishing meerkat running from foraging despite high class imbalance. This was due, in part, to the choice of a sufficiently high sampling frequency of 100 Hz. We found that meerkat running had a characteristic frequency of around 4 Hz, and a good rule-of-thumb is to oversample by about 20 times when using a noisy sensor (Boyd, Sundaram, & Shrivastava, 2010), although the sampling frequency could theoretically be reduced to the Nyquist limit of twice the maximum frequency of interest. Finally, while *stdNorm* and *fftPeakPowerAvg*, used in foraging vs. running classification, make no assumptions about sensor orientation with respect to the animal, *meanSurge*, used in vigilance vs. resting classification, assumes knowledge of sensor placement around the meerkat's neck. Higher running classification accuracy (compared to that for resting) might thus indicate that if the features used are independent of sensor orientation with respect to the animal, classification may be more robust for some species, since sensor fixation will inevitably be slightly different across individuals or species (e.g. collars in mammals and back-packs in birds). Studies on energy expenditure (e.g. Gleiss, Wilson, & Shepard, 2011) also suggested that when accelerometers cannot be accurately placed on the animal, using information from all axes together (Vectorial Dynamic Body Acceleration) may perform better than treating each axis independently (Overall Dynamic Body Acceleration).

Through the use of general biomechanical principles characterising animal movement, our conceptually simple, robust classification method may be applicable across a range of species, with different behaviour labels depending on the species' specific context. The biomechanically defined node-based hierarchical format permitting model adaptation to coarser or finer behavioural resolution makes it apt for meaningful model sharing between studies on a given species. The advantage and novelty of our method is that it allows a high classification performance and, at the same time, a physical and biomechanical interpretation of the classification outcomes. The incorporation of common machine learning algorithms found in all popular, existing packages in `Python`, `MatLab` and `R` makes this method simple and accessible.

ACKNOWLEDGEMENTS

We thank the Northern Cape Conservation Authority for permission to conduct this research, and the farmers neighbouring the KRC for granting us access to their private land. We thank the Trustees of the KRC for access to research facilities in the Kuruman River Reserve and the Directors of the KMP for access to habituated animals with

known life histories. During the period of our project, the long-term research on meerkats was supported by a European Research Council Advanced Grant (No. 294494) to Tim Clutton-Brock, the University of Zurich and the Mammal Research Institute at the University of Pretoria. We thank the field managers, collaborators and assistants for facilitating field work and helping with data collection, and in particular David Gaynor, Tim Vink, Ana Morales González and Héctor Ruiz Villar. Special thanks to Nino Maag for helping with developing the conceptual framework of the project, definition of meerkat behaviours and data collection. We thank Selin Ersoy for annotating all videos, and Pascal Morel for adapting the Physilog IV for this study. This project was supported by the Swiss National Science Foundation (research grant no. CR32I*_159743) to Arpat Ozgul and Kamiar Aminian. We thank the three anonymous reviewers, associate editor and handling editor for their thoughtful comments that helped improve the presentation and clarity of the manuscript.

AUTHORS' CONTRIBUTIONS

K.A. and P.C. developed the research idea, and G.C. and A.O contributed to refinements. P.C. and G.C. supervised the fieldwork. P.C. performed data analyses and led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

DATA ACCESSIBILITY

Data deposited in the Dryad Digital Repository https://doi.org/10.5061/dryad.7q294p8 (Chakravarty et al. 2019).

ORCID

Pritish Chakravarty https://orcid.org/0000-0002-2975-6253
Gabriele Cozzi https://orcid.org/0000-0002-1744-1940
Arpat Ozgul https://orcid.org/0000-0001-7477-2642
Kamiar Aminian https://orcid.org/0000-0002-6582-5375

REFERENCES

Bidder, O. R., Campbell, H. A., Gómez-Laich, A., Urgé, P., Walker, J., & Cai, Y., … Wilson, R. P. (2014). Love thy neighbour: Automatic animal behavioural classification of acceleration data using the k-nearest neighbour algorithm. *PLoS ONE*, *9*(2), e88609. https://doi.org/10.1371/journal.pone.0088609

Boyd, J., Sundaram, H., & Shrivastava, A. (2010). Power-accuracy tradeoffs in human activity transition detection. In *Proceedings of the Conference on Design, Automation and Test in Europe* (pp. 1524–1529). European Design and Automation Association.

Chakravarty, P., Cozzi, G., Ozgul, A., & Aminian, K. (2019). Data from: A novel biomechanical approach for animal behaviour recognition using accelerometers. *Dryad Digital Repository*, https://doi.org/10.5061/dryad.7q294p8

Clutton-Brock, T. H., Gaynor, D., McIlrath, G. M., Maccoll, A. D. C., Kansky, R., Chadwick, P., … Brotherton, P. N. M. (1999). Predation, group size and mortality in a cooperative mongoose, *Suricata suricatta*. *Journal of Animal Ecology*, *68*(4), 672–683. https://doi.org/10.1046/j.1365-2656.1999.00317.x

Cozzi, G., Broekhuis, F., McNutt, J. W., Turnbull, L. A., Macdonald, D. W., & Schmid, B. (2012). Fear of the dark or dinner by moonlight? Reduced temporal partitioning among Africa's large carnivores. *Ecology*, *93*(12), 2590–2599. https://doi.org/10.1890/12-0017.1

Ferraris, F., Grimaldi, U., & Parvis, M. (1995). Procedure for effortless in-field calibration of three-axial rate gyro and accelerometers. *Sensors and Materials*, *7*(5), 311–330.

Gao, L., Campbell, H. A., Bidder, O. R., & Hunter, J. (2013). A Web-based semantic tagging and activity recognition system for species' accelerometry data. *Ecological Informatics*, *13*, 47–56. https://doi.org/10.1016/j.ecoinf.2012.09.003

Gleiss, A. C., Wilson, R. P., & Shepard, E. L. (2011). Making overall dynamic body acceleration work: On the theory of acceleration as a proxy for energy expenditure. *Methods in Ecology and Evolution*, *2*(1), 23–33. https://doi.org/10.1111/j.2041-210X.2010.00057.x

Golabek, K. A., Jordan, N. R., & Clutton-Brock, T. H. (2008). Radiocollars do not affect the survival or foraging behaviour of wild meerkats. *Journal of Zoology*, *274*(3), 248–253. https://doi.org/10.1111/j.1469-7998.2007.00377.x

Goldbogen, J. A., Calambokidis, J., Oleson, E., Potvin, J., Pyenson, N. D., Schorr, G., & Shadwick, R. E. (2011). Mechanics, hydrodynamics and energetics of blue whale lunge feeding: Efficiency dependence on krill density. *Journal of Experimental Biology*, *214*(1), 131–146. https://doi.org/10.1242/jeb.048157

Grünewälder, S., Broekhuis, F., Macdonald, D. W., Wilson, A. M., McNutt, J. W., Shawe-Taylor, J., & Hailes, S. (2012). Movement activity based classification of animal behaviour with an application to data from cheetah (Acinonyx jubatus). *PLoS ONE*, *7*(11), e49120. https://doi.org/10.1371/journal.pone.0049120

Hammond, T. T., Springthorpe, D., Walsh, R. E., & Berg-Kirkpatrick, T. (2016). Using accelerometers to remotely and automatically characterize behavior in small animals. *Journal of Experimental Biology*, *219*(11), 1618–1624. https://doi.org/10.1242/jeb.136135

He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, *21*(9), 1263–1284.

Hodge, S. J., Manica, A., Flower, T. P., & Clutton-Brock, T. H. (2008). Determinants of reproductive success in dominant female meerkats. *Journal of Animal Ecology*, *77*(1), 92–102. https://doi.org/10.1111/j.1365-2656.2007.01318.x

Jordan, N. R., Cherry, M. I., & Manser, M. B. (2007). Latrine distribution and patterns of use by wild meerkats: Implications for territory and mate defence. *Animal Behaviour*, *73*(4), 613–662.

Kawabe, R., Kawano, T., Nakano, N., Yamashita, N., Hiraishi, T.,& Naito, Y. (2003). Simultaneous measurement of swimming speed and tail beat activity of free-swimming rainbow trout Oncorhynchus mykiss using an acceleration data-logger. *Fisheries science*, *69*(5), 959–965.

Laich, A. G., Wilson, R. P., Quintana, F., & Shepard, E. L. (2008). Identification of imperial cormorant Phalacrocorax atriceps behaviour using accelerometers. *Endangered species research*, *10*, 29–37.

Lemaitre, G., Nogueira, F., & Aridas, C. K. (2017). Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. *Journal of Machine Learning Research*, *18*(17), 1–5.

McClune, D. W., Marks, N. J., Wilson, R. P., Houghton, J. D., Montgomery, I. W., & McGowan, N. E., … Scantlebury, M. (2014). Tri-axial accelerometers quantify behaviour in the Eurasian badger (*Meles meles*): Towards an automated interpretation of field data. *Animal Biotelemetry*, *2*(1), 5. https://doi.org/10.1186/2050-3385-2-5

McGregor, S. J., Busa, M. A., Yaggie, J. A., & Bollt, E. M. (2009). High resolution MEMS accelerometers to estimate VO2 and compare running mechanics between highly trained inter-collegiate and untrained runners. *PLoS ONE*, *4*(10), e7355. https://doi.org/10.1371/journal.pone.0007355

Naito, Y., Bornemann, H., Takahashi, A., McIntyre, T., & Plötz, J. (2010). Fine-scale feeding behavior of Weddell seals revealed by a mandible accelerometer. *Polar Science*, *4*(2), 309–316. https://doi.org/10.1016/j.polar.2010.05.009

Nathan, R., Spiegel, O., Fortmann-Roe, S., Harel, R., Wikelski, M., & Getz, W. M. (2012). Using tri-axial acceleration data to identify behavioral modes of free-ranging animals: General concepts and tools illustrated for griffon vultures. *Journal of Experimental Biology*, *215*(6), 986–996. https://doi.org/10.1242/jeb.058602

Pedregosa, , et al. (2011). Scikit-learn: Machine learning in python. *JMLR*, *12*, 2825–2830.

Resheff, Y. S., Rotics, S., Harel, R., Spiegel, O., & Nathan, R. (2014). AcceleRater: A web application for supervised learning of behavioral modes from acceleration measurements. *Movement Ecology*, *2*(1), 27. https://doi.org/10.1186/s40462-014-0027-0

Ropert-Coudert, Y., Kato, A., Wilson, R. P., & Cannell, B. (2006). Foraging strategies and prey encounter rate of free-ranging Little Penguins. *Marine Biology*, *149*(2), 139. https://doi.org/10.1007/s00227-005-0188-x

Ropert-Coudert, Y., & Wilson, R. P. (2005). Trends and perspectives in animal-attached remote sensing. *Frontiers in Ecology and the Environment*, *3*(8), 437–444. https://doi.org/10.1890/1540-9295(2005)003[0437:TAPIAR]2.0.CO;2

Sellers, W. I., Varley, J. S., & Waters, S. S. (1998). Remote monitoring of locomotion using accelerometers: A pilot study. *Folia Primatologica*, *69*(Suppl. 1), 82–85. https://doi.org/10.1159/000052700

Shepard, E. L., Wilson, R. P., Quintana, F., Laich, A. G., Liebsch, N., & Albareda, D. A., … Newman, C. (2008). Identification of animal movement patterns using tri-axial accelerometry. *Endangered Species Research*, *10*, 47–60. https://doi.org/10.3354/esr00084

Soltis, J., Wilson, R. P., Douglas-Hamilton, I., Vollrath, F., King, L. E., & Savage, A. (2012). Accelerometers in collars identify behavioral states in captive African elephants Loxodonta africana. *Endangered Species Research*, *18*(3), 255–263. https://doi.org/10.3354/esr00452

Wang, Y., Nickel, B., Rutishauser, M., Bryce, C. M., Williams, T. M., Elkaim, G., & Wilmers, C. C. (2015). Movement, resting, and attack behaviors of wild pumas are revealed by tri-axial accelerometer measurements. *Movement Ecology*, *3*(1), 2. https://doi.org/10.1186/s40462-015-0030-0

Watanabe, S., Izawa, M., Kato, A., Ropert-Coudert, Y., & Naito, Y. (2005). A new technique for monitoring the detailed behaviour of terrestrial animals: A case study with the domestic cat. *Applied Animal Behaviour Science*, *94*(1), 117–131. https://doi.org/10.1016/j.applanim.2005.01.010

Watanabe, Y. Y., & Takahashi, A. (2013). Linking animal-borne video to accelerometers reveals prey capture variability. *Proceedings of the National Academy of Sciences*, *110*(6), 2199–2204. https://doi.org/10.1073/pnas.1216244110

Wilson, R. P., White, C. R., Quintana, F., Halsey, L. G., Liebsch, N., Martin, G. R., & Butler, P. J. (2006). Moving towards acceleration for estimates of activity-specific metabolic rate in free-living animals: The case of the cormorant. *Journal of Animal Ecology*, *75*(5), 1081–1090. https://doi.org/10.1111/j.1365-2656.2006.01127.x

Yoda, K., Naito, Y., Sato, K., Takahashi, A., Nishikawa, J., & Ropert-Coudert, Y., … Le Maho, Y. (2001). A new technique for monitoring the behaviour of free-ranging Adelie penguins. *Journal of Experimental Biology*, *204*(4), 685–690.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Chakravarty P, Cozzi G, Ozgul A, Aminian K. A novel biomechanical approach for animal behaviour recognition using accelerometers. *Methods Ecol Evol*. 2019;00:1–13. https://doi.org/10.1111/2041-210X.13172